

DIALOG(R) File 351:Derwent WPI  
(c) 2003 Thomson Derwent. All rts. reserv.

014417100 \*\*Image available\*\*  
WPI Acc No: 2002-237803/200229

XRPX Acc No: N02-183070

Speech recognition method involves using search space in which a Hidden Markov model-to-Hidden Markov model transition is not allowed for specific feature vector sequences

Patent Assignee: CANON KK (CANON); YAMAMOTO H (YAMA-I)

Inventor: YAMAMOTO H

Number of Countries: 002 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20010032075	A1	20011018	US 2001817039	A	20010327	200229 B
JP 2001282283	A	20011012	JP 200099536	A	20000331	200229

Priority Applications (No Type Date): JP 200099536 A 20000331

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20010032075	A1	12		G10L-015/14	
JP 2001282283	A	6		G10L-015/14	

Abstract (Basic): US 20010032075 A1

NOVELTY - The speech recognition of input signal is performed using the search space in which Hidden Markov model (HMM)-to-HMM transition is not allowed for the specific feature vector sequences.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (a) Computer program;
- (b) Recording medium which stores the speech recognition program;
- (c) Speech recognition device

USE - For speech recognition using Hidden Markov model (HMM).

ADVANTAGE - High speed speech recognition is realized, thereby reducing the processing amount by creating optimum path quickly.

DESCRIPTION OF DRAWING(S) - The figure shows the flowchart illustrating speech recognition method.

pp; 12 DwgNo 2/7

Title Terms: SPEECH; RECOGNISE; METHOD; SEARCH; SPACE; HIDE; MARKOV; MODEL; HIDE; MARKOV; MODEL; TRANSITION; ALLOW; SPECIFIC; FEATURE; VECTOR; SEQUENCE

Derwent Class: P86; T01; W04

International Patent Class (Main): G10L-015/14

File Segment: EPI; EngPI

Manual Codes (EPI/S-X): T01-C08A; T01-J18; T01-S03; W04-V04A



(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開2001-282283

(P2001-282283A)

(43)公開日 平成13年10月12日 (2001.10.12)

(51)Int.Cl.<sup>7</sup>

G 1 0 L 15/14

識別記号

F I

テ-マコ-ト(参考)

G 1 0 L 3/00

5 3 5 B 5 D 0 1 5

審査請求 未請求 請求項の数11 O L (全 6 頁)

(21)出願番号 特願2000-99536(P2000-99536)

(71)出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(22)出願日 平成12年3月31日 (2000.3.31)

(72)発明者 山本 寛樹

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(74)代理人 100076428

弁理士 大塚 康徳 (外2名)

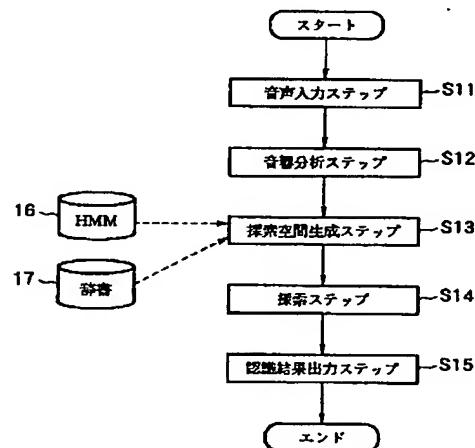
Fターム(参考) 5D015 HH11

(54)【発明の名称】 音声認識方法及び装置と記憶媒体

(57)【要約】

【課題】 ビタビサーチにより最適パスを探索する際に探索すべきパスを減らすことにより、音声認識のための探索処理に要する処理量を削減する。

【解決手段】 音声信号を入力して音声認識を行う音声認識装置における音声認識方法であって、入力された音声信号をA/D変換し、そのデジタル信号を基に特徴ベクトル系列を抽出する(S12)。その特徴ベクトル系列と、予め音声単位ごとに用意されたHMM16とで規定される探索空間であって、特定の特徴ベクトル系列でHMM間の遷移を許可する探索空間を規定し、その空間内で、音声信号に対する音響尤度が最大となる最適パスを探索して認識結果を求める(S14)、その認識結果を出力する(S15)。



【特許請求の範囲】

【請求項1】 入力された音声信号から特徴ベクトル系列を抽出するステップと、特定の特徴ベクトル系列でHMM間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識するステップと、を有することを特徴とする音声認識方法。

【請求項2】 前記探索空間は、1つ以上の認識対象語に対応するHMMの状態系列と前記特徴ベクトル系列との二軸で規定される空間であることを特徴とする請求項1に記載の音声認識方法。

【請求項3】 前記探索空間は、N個の特徴ベクトル系列毎にのみHMM間の遷移を許可することを特徴とする請求項1又は2に記載の音声認識方法。

【請求項4】 前記Nの値は、 $2 \leq N \leq 4$ の範囲にあることを特徴とする請求項3に記載の音声認識方法。

【請求項5】 前記HMMは、音素、音節、単語、diphoneのいずれかに対応することを特徴とする請求項1に記載の音声認識方法。

【請求項6】 音声認識装置を制御するためのプログラムを記憶したコンピュータが読み取り可能な記憶媒体であって、請求項1乃至5のいずれか1項に記載の音声認識方法を計算機を用いて実現する際に、計算機によって実行される音声認識プログラムを記憶したことを特徴とするコンピュータにより読み取り可能な記憶媒体。

【請求項7】 入力された音声信号から特徴ベクトル系列を抽出する抽出手段と、特定の特徴ベクトル系列でHMM間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識する認識手段と、を有することを特徴とする音声認識装置。

【請求項8】 前記探索空間は、1つ以上の認識対象語に対応するHMMの状態系列と前記特徴ベクトル系列との二軸で規定される空間であることを特徴とする請求項7に記載の音声認識装置。

【請求項9】 前記探索空間は、N個の特徴ベクトル系列毎にのみHMM間の遷移を許可することを特徴とする請求項7又は8に記載の音声認識装置。

【請求項10】 前記Nの値は、 $2 \leq N \leq 4$ の範囲にあることを特徴とする請求項9に記載の音声認識装置。

【請求項11】 前記HMMの状態系列は、音素、音節、単語、diphoneのいずれかに対応することを特徴とする請求項7に記載の音声認識装置。

【発明の詳細な説明】

【0001】 【発明の属する技術分野】本発明は、隠れマルコフモデルを用いた音声認識方法及びその装置と、その音声認識を実行するプログラムを記憶したコンピュータにより読み取り可能な記憶媒体に関するものである。

【0002】 【従来の技術】近年、音声認識の有効な方法として、隠れマルコフモデル (Hidden Markov Model、以下、HMMと記す) を用いた方法の研究、応用が進み、多くの音声認識システムなどで用いられている。

【0003】

【発明が解決しようとする課題】図6は、HMMを用いた従来の音声認識方法の一例を示すフローチャートである。

【0004】まずステップS1の音声入力において、マイクロホンなどから入力された音声信号をA/D変換してデジタル信号に変換する。次にステップS2では、ステップS1で変換した音声信号を音響分析し、特徴ベクトルの時系列を抽出する。この音響分析では、時間的に変化する連続波形である音声信号に対して、30ミリ秒程度の窓幅の分析窓を設け、この分析窓を窓幅の $1/3$ ～ $1/2$ 程度 (10ミリ秒～15ミリ秒) ずらしながら音響分析する。各窓内の分析結果は特徴ベクトルとして出力するため、tをフレーム番号とすると、音声信号は特徴ベクトル系列O(t) ( $1 \leq t \leq T$ ) に変換される。

【0005】次にステップS3に進み、所定の構成単位からなるHMMを保持するHMMデータベース5と、認識対象語とHMMの状態系列との対応関係を記述した辞書6を用いて、HMMの状態系列と入力音声の特徴ベクトル系列とを二軸とする探索空間を生成し、この探索空間上で音響尤度が最大となる最適パスをビタビ探索を用いて求める。

【0006】この最適パス探索の詳細な手順を図7を用いて以下で述べる。

【0007】図7は、音素を構成単位としたHMMを用いて、2つの単語『あき』『あき』を連続音声認識する場合の探索空間および探索の様子を表している。この図7において、横軸は特徴ベクトル系列の一例で、縦軸はHMMの状態系列の一例である。

【0008】まず、HMMデータベース5と、認識対象語とHMM5の状態系列との対応関係を記述した辞書6から、1つ以上の認識対象語に対応するHMMの状態系列を生成する。こうして生成されたHMMの状態系列は図7の縦軸のようになる。

【0009】こうして生成されたHMMの状態系列と特徴ベクトル系列とから二次元の格子状の探索空間が形成される。

【0010】次に、図7に示した探索空間上の「STAR T」から出発し「END」に到達する全ての経路(パス)について、各格子点における状態出力確率と各格子点間の遷移に対応するHMMの状態遷移確率とから累積音響尤度が最大となる最適パスを探索する。

【0011】まず、探索空間の各格子点(状態仮説)について、その格子点に到達するまでの累積音響尤度(状態仮説尤度)を  $t = 1$  から  $t = T$  まで順番に求める。第  $t$  フレームの状態  $s$  の状態仮説尤度  $H_{s,t}$  は以下の式で求める。

【0012】

$$H(s, t) = \max_{s' \in S'(s)} H(s', t-1) \times a(s', s) \times b(s, O(t))$$

ここで、 $S'(s)$ は、状態  $s$  に接続する状態の集合、 $a(s', s)$  は、状態  $s'$  から状態  $s$  への遷移確率、 $b(s, O(t))$  は、特徴ベクトル  $O(t)$  に対する状態  $s$  の状態出力確率である。

$$\max_{s \in S_f} H(s, T) \times a(s, s')$$

ここで、 $S_f$  は「END」に到達可能な音素 HMM の状態の集合、即ち、各認識対象語を表す HMM の最終状態の集合を表す。また  $a(s, s')$  は、状態  $s$  から他の状態へ遷移する確率である。

【0014】以上の計算の過程で、各状態仮説の状態仮説尤度を求める際に、状態仮説尤度が最大となる遷移元の状態（式（1）における  $s'$ ）を記憶しておき、これを辿ることで音響尤度が最大となる最適パスが求まる。

【0015】以上の手順で求めた最適パスに対応する HMM の状態系列を求め、さらにその受胎系列に対応する認識対象語を認識結果とする。図7で太線で示したパスが累積音響尤度を最大にする最適パスであった場合、このパスは音素 HMM /a//k//a/ の状態を通るため、この場合の音声認識結果は「あか」となる。最後にステップ S4 に進み、その認識結果を表示装置等に表示したり、或いは他の処理に渡したりする。

【0016】しかしながら、図7に示した探索空間は、認識対象語数、発声時間に比例して大きくなり、このような探索空間の拡大に伴って最適パスの探索処理の処理量が飛躍的に増加する。これにより、大語彙の音声認識を実現する場合や、処理能力が劣る計算機で音声認識を実現する場合に音声認識の応答速度が遅くなるという問題があった。

【0017】本発明は上記従来例に鑑みてなされたもので、音声認識のための探索処理に要する処理量を削減して高速な音声認識を可能にした音声認識方法及びその装置と記憶媒体を提供することを目的とする。

#### 【0018】

【課題を解決するための手段】上記目的を達成するためには本発明の音声認識装置は以下のような構成を備える。即ち、入力された音声信号から特徴ベクトル系列を抽出する抽出手段と、特定の特徴ベクトル系列で HMM 間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識する認識手段と、を有することを特徴とする。

【0019】上記目的を達成するために本発明の音声認識方法は以下のような工程を備える。即ち、入力された音声信号から特徴ベクトル系列を抽出するステップと、特定の特徴ベクトル系列で HMM 間の遷移を許可しない探索空間を用いて、前記音声信号を音声認識するステップと、を有することを特徴とする。

#### 【0020】

【発明の実施の形態】以下、添付図面を参照して本発明

…式（1）

【0013】以上で求めた状態仮説尤度を用いて、「END」に到達する最適パスの音響尤度は以下の式で求められる。

…式（2）

の好適な実施の形態を詳細に説明する。

【0021】図1は本発明の実施の形態に係る音声認識装置のハードウェア構成を示すブロック図である。

【0022】図1において、101は出力部で、例えば表示部や印刷部等を有し、音声認識の結果、或いはその音声認識の結果から得られた応答（文書データ）を出力する。102は入力部で、ここでは例えばマイクロフォンなどの音声を入力するための構成を備えている。またこの入力部102は、キーボードやマウス等のように、オペレータにより操作されて各種データを入力するための構成を備えている。103は中央処理部（CPU）で、数値演算やこの音声認識装置全体の動作制御等を行なう。104は記憶部で、ディスク装置等の外部メモリ装置や、RAM、ROM等の内部メモリを含み、この記憶部には、本実施の形態の手順や処理を実行するための制御プログラム、更にはこの処理に必要な一時的データおよび認識対象語と HMM の対応関係を示した辞書、HMM 等が格納されている。105は音声認識ユニットである。

【0023】以上の構成を備える音声認識ユニット105の動作を以下に詳しく説明する。

【0024】本実施の形態では、HMM の構成単位を音素とし、「あか」「あき」を認識対象語とする、HMM の状態系列を用いて、入力音声を連続音声認識を行う場合について説明する。

【0025】図5は、本実施の形態における認識対象語を示す図で、各認識対象語は音素 HMM で構成されている。また、各音素 HMM は、図4に示すように複数の状態の連結として構成されている。

【0026】図4では3つの状態（S1、S2、S3）が示されており、状態 S1 におけるループ確率は a11、状態 S1 から状態 S2 への遷移確率が a12 で示されている。本実施の形態では、ビタビアルゴリズムを用いて、これら状態間での遷移確率等に基づいて、探索空間内の各経路スコアを求め、そのスコア値の累積値（尤度）が最も大きい経路を探索する。

【0027】図2は、本実施の形態に係る音声合成ユニット105における音声認識処理を示すフローチャートである。なお、図6の HMM を用いた音声認識方法と同様の過程については、詳細な説明を省略し、ステップ S1-1 の音声入力処理（ステップ S1 に対応する）、ステップ S1-2 の音響分析処理（ステップ S2 に対応する）

により、入力音声から特徴ベクトル系列 $O(t)$  ( $1 \leq t \leq T$ ) を抽出した後の探索処理から説明を行う。

【0028】ステップS13の探索空間生成処理において、音素を構成単位とするHMMを保持するHMMデータベース16と、認識対象語とHMMの状態系列との対応関係を記述した辞書17とを用いて、1つ以上の認識対象語に対応するHMMの状態系列を生成し、特徴ベクトル系列 $O(t)$ と、このHMMの状態系列とからなる二軸の探索空間を生成する(図3及び図7参照)。

【0029】図3は、本実施の形態に係る音声認識装置における1方向の探索処理経路を説明する図である。

【0030】図3に示す本実施の形態と、図7の構成との相違点は、図3では、探索空間を生成する際に、特定のフレームのみにHMM間の遷移を許可するパスを用意し、それ以外のフレームではHMM間の遷移を許可しない点にある。これにより、探索空間における最適パスを探索する際の探索すべきパスの数を減少させることができ、これにより処理速度を高めることができる。ここでは、例えば、HMM間の遷移を許すフレームを、例えば $N$  ( $2 \leq N \leq 4$ ) フレーム間隔というように設定する。

【0031】図3の例では、 $N=3$ として、3フレームごとにHMM間での遷移を許すように探索空間が設定されている。この図3と図7とを比較すると、図7の音声認識方法で生成される探索空間に比べ、HMM間を遷移するパスの数が大幅に削減されているのがわかる。

【0032】図3の例では、各音素(/a/, /k/, /i/)のHMMは3つの状態を有し、それぞれ所定の遷移規則に従って他のHMMに遷移する。本実施の形態では、これらHMM間での遷移を $N$  (=3) フレーム毎に許可している。即ち、特徴ベクトル系列 $O(2)$ ,  $O(5)$ ,  $O(8)$ , ...,  $O(T-1)$ でのみHMM間での遷移が許可されている。

【0033】次にステップS14に進み、探索処理において、図3の「START」から出発して「END」に到達する全ての経路(パス)について、各状態仮説の累積尤度における状態出力確率と各格子点間の遷移に対応するHMMの状態遷移確率とから累積音響尤度を計算し、その計算した累積音響尤度が最も大きくなるような最適パスを探索する。なお、この最適パスの探索方法は、図6のステップS3と同様の処理で求まるので、その説明を省略する。

【0034】こうして求めた最適パス上のHMM系列の認識対象語を認識結果とし、ステップS15の認識結果出力処理で、その認識結果を出力部101の表示装置に表示したり、他の処理に渡したりする。

【0035】以上説明したように本実施の形態によれば、特徴ベクトル系列とHMMの状態系列とを用いて探索空間を生成する際に、HMM間での遷移を $N$  フレーム毎にのみ許可することにより、探索するパスの数を減らして、認識処理速度をより高めることができる。

【0036】本実施の形態によれば、図2のステップS13において、HMM間の遷移を許可するフレームを $N$  ( $N=3$ ) フレーム単位とする場合について説明したが、これに限るものではない。例えば、上述の探索空間は、認識対象語の増加や発声時間の増加に伴って拡大することを考慮し、認識対象語の増加または発声時間の増加に応じて、HMM間の遷移を許可するフレームの間隔を $2 \leq N \leq 4$  の範囲において段階的に広げることも可能である。また、認識対象語の増加と発声時間の増加の双方を考慮して、フレームの間隔を $2 \leq N \leq 4$  の範囲において段階的に変更することも可能である。このように構成することにより、探索空間の規模に応じて適応的に探索パスの削減を行え、認識処理の速度を高めることができる。

【0037】また、本実施の形態のステップS13では、HMMの状態系列内に存在する全てのHMMに対して、HMM間の遷移を許可するフレームを $N$  ( $N=3$ ) フレーム単位とする場合について説明したが、これに限るものではない。例えば、HMM間の遷移を許可するフレームの間隔を、所定のHMM間において $2 \leq N \leq 4$  の範囲で変更することも可能である。また、所定数の特徴ベクトル系列ごとに、フレーム間隔を可変とすることも可能である。これにより、他のHMMへ遷移する頻度が高いHMMと、他のHMMへ遷移する頻度が低いHMMとでフレーム間隔を変更することができる。このように構成することにより、認識率の向上と探索空間の縮小とを同時に実現することができる。

【0038】また、本実施の形態では、HMMの構成単位を音素として説明したが、これに限るものではない。音節、単語、diphone等の音韻を構成単位としてもよい。

【0039】また本実施の形態では、日本語の単語を認識する例について説明したが、これに限るものではない。日本語以外の言語にも適用できる。

【0040】なお本発明は、複数の機器(例えばホストコンピュータ、インターフェース機器、リーダ、プリンタなど)から構成されるシステムに適用しても、一つの機器からなる装置(例えば、複写機、ファクシミリ装置など)に適用してもよい。

【0041】また、本発明の目的は、前述した実施の形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体(または記録媒体)を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ(またはCPUやMPU)が記憶媒体に格納されたプログラムコードを読み出し実行することによって達成される。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。また、コンピュータが読み出したプログラムコードを実行することにより、前述

した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているオペレーティングシステム（OS）などが実際の処理の一部または全部を行い、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【0042】さらに、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張カードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張カードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【0043】以上説明したように本実施の形態によれば、最尤状態系列を探索する探索空間を形成する際に、特定のフレームのみHMM間の遷移を許すことによって、探索すべきバスを削減し、最適バスの探索に要する処理量を削減できる。この結果、従来の方法よりも高速な音声認識を実現することが可能となる。

【0044】

【発明の効果】以上説明したように本発明によれば、最適バスを探索する際の探索すべきバスを減らすことができ、音声認識のための探索処理に要する処理量を削減して高速な音声認識を可能にできるという効果がある。

【図面の簡単な説明】

【図1】本発明の実施の形態に係る音声認識装置のハードウェア構成を示すブロック図である。

【図2】本発明の実施の形態に係る音声認識装置における音声認識処理手順を示すフローチャートである。

【図3】本発明の実施の形態に係る探索処理を行う経路を説明する図である。

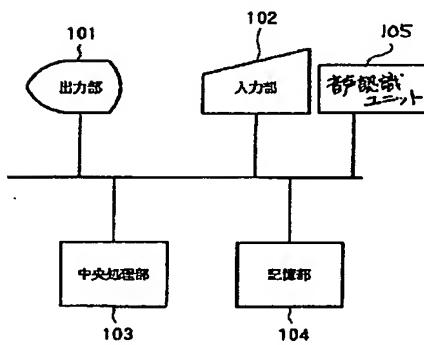
【図4】隠れマルコフモデルを説明する図である。

【図5】本発明の実施の形態における、認識対象語が複数の音素モデルで構成されている様子を示した図である。

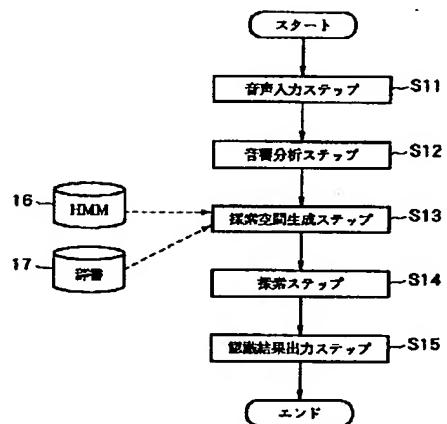
【図6】従来の音声認識処理の処理手順を示したフローチャートである。

【図7】従来の音声認識方法における探索処理経路を説明する図である。

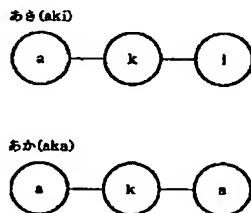
【図1】



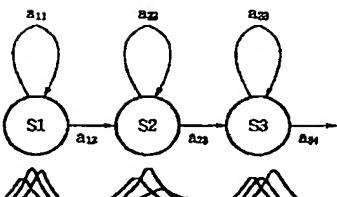
【図2】



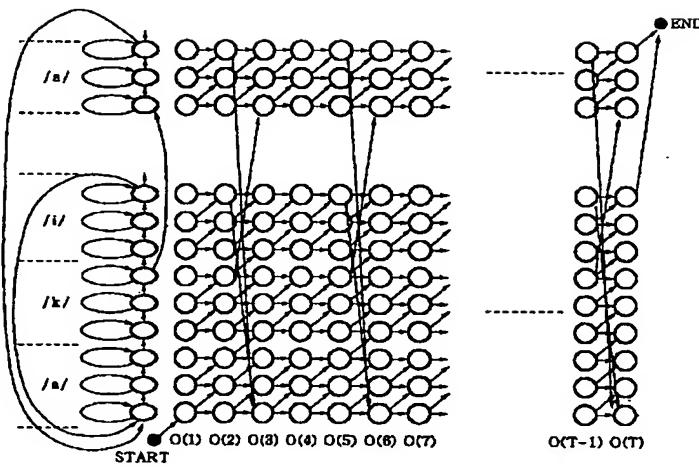
【図5】



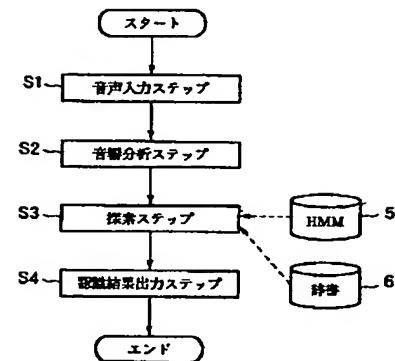
【図4】



【図3】



【図6】



【図7】

